**Capstone 2:**

Project Proposal

Data: [KDD Cyberattack | Kaggle](https://www.kaggle.com/datasets/slashtea/kdd-cyberattack).

Network intrusion detection software is a valuable tool to stop unauthorized users from getting access to a network. The purpose of this project is to build a predictive model that can discern between bad connections (intrusions or attacks) and normal connections. This model will use data from the 1998 DARPA Intrusion Detection Evaluation Program that was prepared and managed by MIT’s Lincoln Lab. For nine weeks the researchers acquired raw TCP dump data peppering the data with bad connections. This environment simulated a typical US Air Force LAN.

This is a community research project, so the stakeholders are more abstract. The work that this project could accomplish would provide a tool to fight against cyber-attacks by understanding the signals that precede the attacks. With a working model, companies and organizations could implement better security practices by adjusting their networks to align with the suggestions that the model proposed.

The biggest limitation of the data is that an attack is a binary event even though there are a diverse set of attack types. It is impossible to try to come up with a model that would be able to predict an unauthorized user while also trying to predict a denial of service. That is why the aim of this project is to parse through the distinct kinds of attacks, sort them into groups that are similar to each other and come up with more interesting analysis.

The goal of this project is to create a prediction model that has a level of precision that is greater than random chance with a statistically significant p-value. This model may only have a subset of the attacks that can correspond to the independent variables, but it will still provide insight into limiting costly intrusions.